EAST Search History

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S97	149	(ocr$3 with text$1) and (keyword$1 with list$1)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/07 12:14
S98	54	(ocr$3 with text$1) and (keyword$1 with list$1) and dictionar$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/07 12:18
S99	8	(ocr$3 with text$1) and (keyword$1 with list$1) and dictionar$3 with keyword$1	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/07 12:15
S100	2	(ocr$3 with text$1) and (keyword$1 with list$1) and (dictionar$3 with (keyword$1 and phrase$1))	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/07 12:15
S101	91	(ocr$3) and (keyword$1 with list$1) and dictionar$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/07 12:18
S102	87	(ocr$3) and (keyword$1 with list$1) and dictionar$3 and search	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/07 12:18
S103	89	(ocr$3) and (keyword$1 with list$1) and dictionar$3 and search$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 16:41
S104	208	(ocr$3) and (keyword$1 with list$1) and (dictionar$3 with keyword$1 or phrase$1 or term$1) and (search$3 with keyword$1 or phrase$1 or term$1)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/07 12:20

EAST Search History

S10 5	58	(ocr$3) and (keyword$1 with list$1) and (dictionar$3 with (keyword$1 or phrase$1 or term$1)) and (search$3 with (keyword$1 or phrase$1 or term$1))	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/07 12:20
S10 6	36	(ocr$3) and (keyword$1 with list$1) and (dictionar$3 with (keyword$1 or phrase$1 or term$1)) and (search$3 with (keyword$1 or phrase$1 or term$1)) and (frequenc$3 with (keyword$1 or phrase$1 or term$1))	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/07 12:21
S10 7	1682255	(ocr$3) with document$1	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:11
S10 8	184	(ocr$3) with document$1 and ((extract$3 or identif$3) with keyword$1)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:10
S10 9	15	(ocr$3) with document$1 and ((extract$3 or identif$3) with keyword$1) and (keyword$1 with dictionar$3)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 16:44
S11 0	13	(ocr$3) with document$1 and ((extract$3 or identif$3) with keyword$1) and (keyword$1 with dictionar$3) and languag$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 16:54
S11 1	12	(ocr$3) with document$1 and ((extract$3 or identif$3) with keyword$1) and (keyword$1 with dictionar$3) and languag$3 and search$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:15
S11 2	27	(ocr$3) with document$1 and search$3 and (xml with list$3)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:07

EAST Search History

S11 3	2645	search$3 and (xml with list$3)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:07
S11 4	285	search$3 and (xml with list$3) same output$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:08
S11 5	5	search$3 and (xml with list$3) with (metadata) same output$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:09
S11 6	2	search$3 and (xml with list$3) with (distance$1) same output$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:09
S11 7	5	search$3 and (xml with list$3) with (status$2) same output$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:09
S11 8	7	(ocr$3) with document$1 and ((extract$3 or identif$3) with keyword$1) and (xml with list$3)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:11
S11 9	31	(ocr$3) with document$1 and (xml with list$3)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:11
S12 0	4	(ocr$3) with document$1 and (xml with list$3) same output	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:11

S12 2	4	(ocr$3) with document$1 and (xml with list$3) same output$4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:12
S12 3	352	((extract$3 or identif$3) with keyword$1) and (keyword$1 with dictionar$3) and languag$3 and search$3	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:15
S12 4	9	((extract$3 or identif$3) with keyword$1) and (keyword$1 with dictionar$3) and languag$3 and search$3 and (xml with list$3) same output$4	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/17 17:15
S12 5	1	("20050086205").PN.	US-PGPUB; USPAT; USOCR	OR	OFF	2008/01/18 14:03
S12 6	11162	((707/3) or (707/5)).CCLS.	US-PGPUB; USPAT; USOCR	OR	OFF	2008/01/18 14:04
S12 7	184	(ocr$3) with document$1 and ((extract$3 or identif$3) with keyword$1)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/18 14:04
S12 8	11	S126 and S127	USPAT	OR	OFF	2008/01/18 14:04
S12 9	45160	((707/1-10) or (707/100-102) or (707/200-206)).CCLS.	US-PGPUB; USPAT; USOCR	OR	OFF	2008/01/18 14:06
S13 0	15	(ocr$3) with document$1 and ((extract$3 or identif$3) with keyword$1) and (keyword$1 with dictionar$3)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/18 14:06
S13 1	2	S129 and S130	USPAT	OR	OFF	2008/01/18 14:06
S13 2	3981	(707/104.1).CCLS.	USPAT; USOCR	OR	OFF	2008/01/18 16:39

EAST Search History

S13 3	1682255	(ocr$3) with document$1	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2008/01/18 16:39
S13 4	27	S132 and S133	USPAT	OR	OFF	2008/01/18 16:39

P@RTAL

USPTO

Search: ◉ The ACM Digital Library   ○ The Guide

OCR and dictionary and search

SEARCH

THE ACM DIGITAL LIBRARY

⁍ Feedback

OCR and dictionary and search
Terms used: **OCR**   **dictionary**   **search**

Found **96** of **238,048**

| Sort results by | relevance ▼ | 📘 Save results to a Binder | Refine these results with Advanced Search |
| Display results | expanded form ▼ | ☐ Open results in a new window | Try this search in The ACM Guide |

Results 1 - 20 of 96   Result page: **1**   2   3   4   5   next   >>

1   Error correction vs. query garbling for Arabic OCR document retrieval

Kareem Darwish, Walid Magdy
November 2007 **ACM Transactions on Information Systems (TOIS),**
Volume 26 Issue 1
Publisher: ACM

Full text available: 📄 pdf(585.25 KB)   Additional Information: full citation, abstract, references, index terms

Due to the existence of large numbers of legacy documents (such as old books and newspapers), improving retrieval effectiveness for OCR'ed documents continues to be an important problem. This article compares the effect of OCR error correction with and ...

Keywords: Arabic Retrieval, OCR Correction, OCR Retrieval

2   Error correction in a Chinese OCR test collection

Yuen-Hsien Tseng
August 2002 **SIGIR '02:** Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval
Publisher: ACM

Full text available: 📄 pdf(126.00 KB)   Additional Information: full citation, abstract, references, index terms

This article proposes a technique for correcting Chinese OCR errors to support retrieval of scanned documents. The technique uses a completely automatic technique (no manually constructed lexicons or confusion resources) to identify both keywords and ...

Keywords: Chinese, confusing pair, error correction, term clustering

3   A filter based post-OCR accuracy boost system

Eugene Borovikov, Ilya Zavorin, Mark Turner
November 2004 **HDP '04:** Proceedings of the 1st ACM workshop on Hardcopy document processing
Publisher: ACM

Full text available: 📄 pdf(126.24 KB)   Additional Information: full citation, abstract, references, index terms

Our current research effort aims at building a filter based post-OCR accuracy boost system that will combine different post-OCR correction

filters to improve the OCR accuracy better than each individual filter can. In this paper we focus on a Hidden ...

Keywords: HMM, OCR, error correction, information retrieval, machine translation, statistical language models

---

4 Information access in the presence of OCR errors
Kazem Taghva, Thomas Nartker, Julie Borsack
November 2004 **HDP '04:** Proceedings of the 1st ACM workshop on
                        Hardcopy document processing
Publisher: ACM

Full text available: pdf(139.50 KB)   Additional Information: full citation, abstract,
                                                              references, index terms

Over the last 15 years, the Information Science Research Institute (ISRI) at the University of Nevada, Las Vegas (UNLV) has conducted information access research in the presence of OCR errors. Our research has focused on issues associated with the construction ...

Keywords: categorization, document conversion, information extraction, markup

---

5 Adaptive Hindi OCR using generalized Hausdorff image comparison
Huanfeng Ma, David Doermann
September 2003 **ACM Transactions on Asian Language Information
                        Processing (TALIP),** Volume 2 Issue 3
Publisher: ACM

Full text available: pdf(280.45 KB)   Additional Information: full citation, abstract,
                                                              references, index terms

We present an adaptive Hindi OCR implemented as part of a rapidly retargetable language tool effort. The system includes: script identification, character segmentation, training sample creation, and character recognition. In script identification, Hindi ...

Keywords: Optical character recognition (OCR), document processing, generalized Hausdorff image comparison, script identification

---

6 Securing passwords against dictionary attacks
Benny Pinkas, Tomas Sander
November 2002 **CCS '02:** Proceedings of the 9th ACM conference on
                        Computer and communications security
Publisher: ACM

                                    Additional Information: full citation, abstract,
Full text available: pdf(216.72 KB)   references, cited by, index
                                                             terms

The use of passwords is a major point of vulnerability in computer security, as passwords are often easy to guess by automated programs running dictionary attacks. Passwords remain the most widely used authentication method despite their well-known security ...

---

7 Fast Approximate Search in Large Dictionaries
Stoyan Mihov, Klaus U. Schulz
December 2004 **Computational Linguistics,** Volume 30 Issue 4

Publisher: MIT Press

Full text available: pdf(509.51 KB)   Additional Information: full citation, abstract, references, index terms

The need to correct garbled strings arises in many areas of natural language processing. If a dictionary is available that covers all possible input tokens, a natural set of candidates for correcting an erroneous input P is the set of all words ...

8 Joint categorization of queries and clips for web-based video search
Ruofei Zhang, Ramesh Sarukkai, Jyh-Herng Chow, Wei Dai, Zhongfei Zhang
October 2006 **MIR '06:** Proceedings of the 8th ACM international workshop on Multimedia information retrieval
Publisher: ACM

Full text available: pdf(378.76 KB)   Additional Information: full citation, abstract, references, index terms

Building a video search engine on the Web is a very challenging problem. Compared with web page search, video search has its unique characteristics (such as high volume of data for each video, existence of multi-modal information including meta-data, ...

Keywords: experiment, multi-modality based categorization, query categorization, video categorization, web-based video search

9 Multi-modal information retrieval from broadcast video using OCR and speech recognition
Alexander G. Hauptmann, Rong Jin, Tobun Dorbin Ng
July 2002 **JCDL '02:** Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries
Publisher: ACM

Full text available: pdf(205.98 KB)   Additional Information: full citation, abstract, references, cited by, index terms

We examine multi-modal information retrieval from broadcast video where text can be read on the screen through OCR and speech recognition can be performed on the audio track. OCR and speech recognition are compared on the 2001 TREC Video Retrieval evaluation ...

Keywords: multi-modal video information retrieval, optical character recognition OCR, speech recognition

10 A generative probabilistic OCR model for NLP applications
Okan Kolak, William Byrne, Philip Resnik
May 2003 **NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1**, Volume 1
Publisher: Association for Computational Linguistics

Full text available: pdf(144.33 KB)   Additional Information: full citation, abstract, references, cited by

In this paper, we introduce a generative probabilistic optical character recognition (OCR) model that describes an end-to-end process in the noisy channel framework, progressing from generation of true text

through its transformation into the noisy output ...

11 Information retrieval and OCR: from converting content to grasping meaning

Jamie Callan, Paul Kantor, David Grossman
September 2002 **ACM SIGIR Forum**, Volume 36 Issue 2
Publisher: ACM

Full text available: pdf(33.44 KB) Additional Information: full citation, abstract

IR and OCR have largely developed independent standards and metrics, with OCR focused on literal accuracy, and IR focused on essential "content/meaning". With more and more media not only paper, but in multiple image formats, the opportunities and challenges ...

12 Adaptive text correction with Web-crawled domain-dependent dictionaries

Christoph Ringlstetter, Klaus U. Schulz, Stoyan Mihov
October 2007 **ACM Transactions on Speech and Language Processing (TSLP)**, Volume 4 Issue 4
Publisher: ACM

Full text available: pdf(390.15 KB) Additional Information: full citation, abstract, references, index terms

For the success of lexical text correction, high coverage of the underlying background dictionary is crucial. Still, most correction tools are built on top of static dictionaries that represent fixed collections of expressions of a given language. When ...

Keywords: Adaptive techniques, Web crawling, dictionaries, domains, error correction

13 Computer programs for detecting and correcting spelling errors

James L. Peterson
December 1980 **Communications of the ACM**, Volume 23 Issue 12
Publisher: ACM

Full text available: pdf(1.25 MB) Additional Information: full citation, abstract, references, cited by

With the increase in word and text processing computer systems, programs which check and correct spelling will become more and more common. Peterson investigates the basic structure of several such existing programs and their approaches to solving the ...

Keywords: spelling, spelling correction, spelling dictionary, spelling programs

14 Dynamic spatial approximation trees

Gonzalo Navarro, Nora Reyes
August 2007 **Journal of Experimental Algorithmics (JEA)**, Volume 12
Publisher: ACM

Full text available: pdf(2.33 MB) Additional Information: full citation, abstract, references, index terms

Metric space searching is an emerging technique to address the problem

of efficient similarity searching in many applications, including multimedia databases and other repositories handling complex objects. Although promising, the metric space approach ...

Keywords: Multimedia databases, similarity or proximity search, spatial and multidimensional search, spatial approximation tree

15 A new generation of textual corpora: mining corpora from very large collections

Gordon Stewart, Gregory Crane, Alison Babeu
June 2007 **JCDL '07:** Proceedings of the 2007 conference on Digital libraries
Publisher: ACM

Full text available: .pdf(253.35 KB)   Additional Information: full citation, abstract, references, index terms

While digital libraries based on page images and automatically generated text have made possible massive projects such as the Million Book Library, Open Content Alliance, Google, and others, humanists still depend upon textual corpora expensively produced ...

Keywords: OCR evaluation, ancient greek, text alignment

16 A hash code method for detecting and correcting spelling errors

M. Mor, A. S. Fraenkel
December 1982 **Communications of the ACM**, Volume 25 Issue 12
Publisher: ACM

Additional Information: full citation, abstract, references, cited by, index terms

Full text available: .pdf(458.71 KB)

The most common spelling errors are one extra letter, one missing letter, one wrong letter, or the transposition of two letters. Deletion, exchange, and rotation operators are defined which detect and "mend" such spelling errors and thus ...

Keywords: deletion, dictionary, exchange, rotation, spelling, spelling errors

17 Probabilistic structured query methods

Kareem Darwish, Douglas W. Oard
July 2003 **SIGIR '03:** Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval
Publisher: ACM

Additional Information: full citation, abstract, references, cited by, index terms

Full text available: .pdf(301.72 KB)

Structured methods for query term replacement rely on separate estimates of term tes of replacement probabilities. Statistically significantfrequency and document frequency to compute a weight for each query term. This paper reviews prior work on structured ...

Keywords: CLIR, OCR, arabic, structured queries, term replacement

18

Robust document image understanding technologies

Henry S. Baird, Daniel Lopresti, Brian D. Davison, William M. Pottenger
November 2004 **HDP '04:** Proceedings of the 1st ACM workshop on
Hardcopy document processing
Publisher: ACM

Full text available: pdf(92.22 KB)

Additional Information: full citation, abstract, references, cited by, index terms

No existing document image understanding technology, whether experimental or commercially available, can guarantee high accuracy across the full range of documents of interest to industrial and government agency users. Ideally, users should be able to ...

Keywords: OCR error management, document analysis, information retrieval

19 Technique for automatically correcting words in text

Karen Kukich
December 1992 **ACM Computing Surveys (CSUR)**,   Volume 24 Issue 4
Publisher: ACM

Full text available: pdf(6.23 MB)

Additional Information: full citation, abstract, references, cited by, index terms, review

Research aimed at correcting words in text has focused on three progressively more difficult problems:(1) nonword error detection; (2) isolated-word error correction; and (3) context-dependent work correction. In response to the first problem, efficient ...

Keywords: n-gram analysis, Optical Character Recognition (OCR), context-dependent spelling correction, grammar checking, natural-language-processing models, neural net classifiers, spell checking, spelling error detection, spelling error patterns, statistical-language models, word recognition and correction

20 The lifecycle of a digital historical document: structure and content

A. Antonacopoulos, D. Karatzas, H. Krawczyk, B. Wiszniewski
October 2004 **DocEng '04:** Proceedings of the 2004 ACM symposium on
Document engineering
Publisher: ACM

Full text available: pdf(313.24 KB)

Additional Information: full citation, abstract, references, cited by, index terms, review

This paper describes the lifecycle of a digital historical document, from template-based structure definition through to content extraction from the scanned pages and its final reconstitution as an electronic document (combining content and semantic ...

Keywords: digital libraries, document analysis, document architecture, document engineering, historical documents, text enhancement

Results 1 - 20 of 96   Result page: **1**   2   3   4   5   next   >>

PORTAL

USPTO

Search: ⦿ The ACM Digital Library   ⦾ The Guide

search documents and ocr and translate keywords   **SEARCH**

THE ACM DIGITAL LIBRARY

↑ Feedback

search documents and ocr and translate keywords
Terms used: **search documents ocr translate keywords**

Found **23 of 238,048**

Sort results by | relevance ▾ | ● Save results to a Binder

Display results | expanded form ▾ | ☐ Open results in a new window

Refine these results with Advanced Search
Try this search in The ACM Guide

Results 1 - 20 of 23   Result page: **1**   2   next   >>

1 Document digitization lifecycle for complex magazine collection
Sherif Yacoub, John Burns, Paolo Faraboschi, Daniel Ortega, Jose Abad Peiro, Vinay Saxena
November 2005 **DocEng '05**: Proceedings of the 2005 ACM symposium on Document engineering
Publisher: ACM

Additional Information: full citation, abstract, references, cited by, index terms

Full text available: 🖹 pdf(540.79 KB)

The conversion of large collections of documents from paper to digital formats that are suitable for electronic archival is a complex multi-phase process. The creation of good quality images from paper documents is just one phase. To extract relevant ...

Keywords: document analysis and understanding, document digitization, document engineering, preservation of historical content

2 Eclipse modeling framework for document management
Neil Boyette, Vikas Krishna, Savitha Srinivasan
November 2005 **DocEng '05**: Proceedings of the 2005 ACM symposium on Document engineering
Publisher: ACM

Additional Information: full citation, abstract, references, index terms

Full text available: 🖹 pdf(180.43 KB)

The lifecycle of document management applications typically comprises a set of loosely coupled subsystems that provide capture, index, search, workflow, fulfillment and archival features. However, there exists no standard model for composing these elements ...

Keywords: DMS, EMF, IT, ROI, business transformation, documents, eclipse, framework, logistics, modeling, process

3 Extracting mathematical expressions from postscript documents
Michael Yang, Richard Fateman
July 2004 **ISSAC '04**: Proceedings of the 2004 international symposium on Symbolic and algebraic computation
Publisher: ACM

Full text available: 🖹 pdf(137.43 KB)   Additional Information: full citation, abstract,

references, index terms

Full-text indexing of documents containing mathematics cannot be considered a complete success unless the mathematics symbolism is extracted and represented in a standardized form permitting both searching for formulas, and re-use of this information ...

Keywords: digital library, document image analysis, mathematics, optical character recognition, postscript documents

4 Joint visual-text modeling for automatic retrieval of multimedia documents
G. Iyengar, P. Duygulu, S. Feng, P. Ircing, S. P. Khudanpur, D. Klakow, M. R. Krause, R. Manmatha, H. J. Nock, D. Petkova, B. Pytlik, P. Virga
November 2005 **MULTIMEDIA '05**: Proceedings of the 13th annual ACM international conference on Multimedia
Publisher: ACM

Additional Information: full citation, abstract, references, cited by, index terms

Full text available: pdf(435.67 KB)

In this paper we describe a novel approach for jointly modeling the text and the visual components of multimedia documents for the purpose of information retrieval(IR). We propose a novel framework where individual components are developed to model different ...

Keywords: TRECVID, joint visual-text models, multimedia retrieval models, retrieval models

5 Scoring missing terms in information retrieval tasks
Egidio Terra, Charles L.A. Clarke
November 2004 **CIKM '04**: Proceedings of the thirteenth ACM international conference on Information and knowledge management
Publisher: ACM

Additional Information: full citation, abstract, references, cited by, index terms

Full text available: pdf(286.12 KB)

An usual approach to address mismatching vocabulary problem is to augment the original query using dictionaries and other lexical resources and/or by looking at pseudo-relevant documents. Either way, terms are added to form a new query that will be used ...

Keywords: automatic query expansion, document retrieval, passage retrieval

6 The multivalent browser: a platform for new ideas
Thomas A. Phelps, Robert Wilensky
November 2001 **DocEng '01**: Proceedings of the 2001 ACM Symposium on Document engineering
Publisher: ACM

Additional Information: full citation, abstract, references, cited by, index terms

Full text available: pdf(188.51 KB)

The Multivalent Browser is built on a architecture that separates functionality from concrete document format. Almost all functionality is made available via relatively small modules of code called behaviors that programmers can write to extend the core ...

Keywords: annotation, architecture, digital, document, multivalent behavior, paper, scanned

7 The digital atheneum: new approaches for preserving, restoring and analyzing damaged manuscripts

Michael S. Brown, W. Brent

January 2001 **JCDL '01:** Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries

Publisher: ACM

Full text available: .pdf(2.30 MB)   Additional Information: full citation, abstract, references, cited by, index terms

This paper presents research focused on developing new techniques and algorithms for the digital acquisition, restoration, and study of damaged manuscripts. We present results from an acquisition effort in partnership with the British Library, funded ...

Keywords: digital libraries, digital preservation, document analysis, humanities computing, restoration

8 Task-based interaction with an integrated multilingual, multimedia information system: a formative evaluation

Pengyi Zhang, Lynne Plettenberg, Judith L. Klavans, Douglas W. Oard, Dagobert Soergel

June 2007 **JCDL '07:** Proceedings of the 2007 conference on Digital libraries

Publisher: ACM

Full text available: .pdf(720.25 KB)   Additional Information: full citation, abstract, references, index terms

This paper describes a formative evaluation of an integrated multilingual, multimedia information system, a series of user studies designed to guide system development. The system includes automatic speech recognition for English, Chinese, and Arabic, ...

Keywords: cross-language information retrieval, multimedia, user studies

9 Adaptive information extraction

Jordi Turmo, Alicia Ageno, Neus Català

July 2006 **ACM Computing Surveys (CSUR)**, Volume 38 Issue 2

Publisher: ACM

Full text available: .pdf(986.35 KB)   Additional Information: full citation, abstract, references, index terms

The growing availability of online textual sources and the potential number of applications of knowledge acquisition from textual data has lead to an increase in Information Extraction (IE) research. Some examples of these applications are the generation ...

Keywords: Information extraction, machine learning

10 An automatic sign recognition and translation system

Jie Yang, Jiang Gao, Ying Zhang, Xilin Chen, Alex Waibel

November 2001 **PUI '01:** Proceedings of the 2001 workshop on Perceptive user interfaces

Publisher: ACM

Full text available: 📄 pdf(1.21 MB) Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>

A sign is something that suggests the presence of a fact, condition, or quality. Signs are everywhere in our lives. They make our lives easier when we are familiar with them. But sometimes they pose problems. For example, a tourist might not be able ...

Keywords: perceptive user interface, sign detection, sign translation, vision-based interface

---

11 <u>Efficient web browsing on handheld devices using page and form summarization</u>

January 2002 **ACM Transactions on Information Systems (TOIS)**, Volume 20 Issue 1

Publisher: ACM

Full text available: 📄 pdf(4.47 MB) Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>cited by</u>, <u>index terms</u>, <u>review</u>

We present a design and implementation for displaying and manipulating HTML pages on small handheld devices such as personal digital assistants (PDAs), or cellular phones. We introduce methods for summarizing parts of Web pages and HTML forms. Each Web ...

Keywords: PDA, Personal digital assistant, WAP, WML, forms, handheld computers, mobile computing, summarization, ubiquitous computing, wireless computing

---

12 <u>Machine learning in automated text categorization</u>

Fabrizio Sebastiani

March 2002 **ACM Computing Surveys (CSUR)**, Volume 34 Issue 1

Publisher: ACM

Full text available: 📄 pdf(524.41 KB) Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>cited by</u>, <u>index terms</u>

The automated categorization (or classification) of texts into predefined categories has witnessed a booming interest in the last 10 years, due to the increased availability of documents in digital form and the ensuing need to organize them. In the research ...

Keywords: Machine learning, text categorization, text classification

---

13 <u>Towards automatic sign translation</u>

Jie Yang, Jiang Gao, Ying Zhang, Alex Waibel

March 2001 **HLT '01:** Proceedings of the first international conference on Human language technology research

Publisher: Association for Computational Linguistics

Full text available: 📄 pdf(208.10 KB) Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>cited by</u>

Signs are everywhere in our lives. They make our lives easier when we are familiar with them. But sometimes they also pose problems. For example, a tourist might not be able to understand signs in a foreign country. In this paper, we present our efforts ...

Keywords: sign, sign detection, sign recognition, sign translation

14 Commercial applications of natural language processing

Kenneth W. Church, Lisa F. Rau

November 1995 **Communications of the ACM**,   Volume 38 Issue 11

Publisher: ACM

Full text available: pdf(314.22 KB)

Additional Information: full citation, abstract, references, cited by, index terms

Vast quantities of text are becoming available in electronic form, ranging from published documents (e.g., electronic dictionaries, encyclopedias, libraries and archives for information retrieval services), to private databases (e.g., marketing information, ...

15 The use of topic evolution to help users browse and find answers in news video corpus

Shi-Yong Neo, Yuanyuan Ran, Hai-Kiat Goh, Yantao Zheng, Tat-Seng Chua, Jintao Li

September 2007 **MULTIMEDIA '07:** Proceedings of the 15th international conference on Multimedia

Publisher: ACM

Full text available: pdf(626.56 KB)

Additional Information: full citation, abstract, references, index terms

Earlier research in news video has been focusing mainly on improving retrieval accuracies given the limited amount of extractable video semantics. In this paper, we propose an enhancement to news video searching by leveraging extractable video semantics ...

Keywords: event evolution, video analysis, video question answering

16 Towards to an automatic semantic annotation for multimedia learning objects

Stephan Repp, Serge Linckels, Christoph Meinel

September 2007 **Emme '07:** Proceedings of the international workshop on Educational multimedia and multimedia education

Publisher: ACM

Full text available: pdf(381.14 KB)

Additional Information: full citation, abstract, references, index terms

The number of digital video recordings has increased dramatically. The idea of recording lectures, speeches, and other academic events is not new. But, the accessibility and traceability of its content for further use is rather limited. Searching multimedia ...

Keywords: multimedia knowledge base, multimedia retrieval, speech

17 Efficient Web form entry on PDAs

Oliver Kaljuvee, Orkut Buyukkokten, Hector Garcia-Molina, Andreas Paepcke

April 2001 **WWW '01:** Proceedings of the 10th international conference on World Wide Web

Publisher: ACM

Full text available: pdf(398.94 KB)

Additional Information: full citation, references, cited by, index terms

Keywords: PDA, WAP, forms, mobile computing, wireless access

18 PicturePiper: using a re-configurable pipeline to find images on the Web

Adam M. Fass, Eric A. Bier, Eyton Adar

November 2000 **UIST '00:** Proceedings of the 13th annual ACM symposium on User interface software and technology

Publisher: ACM

Full text available: pdf(364.31 KB)   Additional Information: full citation, references, index terms

Keywords: WWW searching, dataflow, image retrieval, pipeline

19 A retrospective look at Greenstone: lessons from the first decade

Ian H. Witten, David Bainbridge

June 2007 **JCDL '07:** Proceedings of the 2007 conference on Digital libraries

Publisher: ACM

Full text available: pdf(572.61 KB)   Additional Information: full citation, abstract, references, index terms

The Greenstone Digital Library Software has helped spread the practical impact of digital library technology throughout the world, with particular emphasis on developing countries. As Greenstone enters its second decade, this article takes a retrospective ...

Keywords: architecture, greenstone, internationalization

20 Adapting handwriting recognition for applications in algebra learning

Lisa Anthony, Jie Yang, Kenneth R. Koedinger

September 2007 **Emme '07:** Proceedings of the international workshop on Educational multimedia and multimedia education

Publisher: ACM

Full text available: pdf(308.71 KB)   Additional Information: full citation, abstract, references, index terms

In this paper we report the progress of our ongoing project exploring the adaptation of handwriting recognition-based interfaces for applications in intelligent tutoring systems for students learning algebra equation-solving. The research is motivated ...

Keywords: handwriting input, handwriting recognition, mathematics learning, recognition accuracy evaluation

Results 1 - 20 of 23   Result page: **1**   2   next   >>

PORTAL

USPTO

Search: ⊙ The ACM Digital Library   ○ The Guide

input documents and ocr and output XML   **SEARCH**

THE ACM DIGITAL LIBRARY   ↑ Feedback

input documents and ocr and output XML
Terms used: **input documents ocr output XML**

Found **38** of **238,048**

Sort results by   relevance ▾   ❤ Save results to a Binder

Display results   expanded form ▾   ☐ Open results in a new window

Refine these results with Advanced Search
Try this search in The ACM Guide

Results 1 - 20 of 38   Result page: **1**   2   next   >>

1   The lifecycle of a digital historical document: structure and content
A. Antonacopoulos, D. Karatzas, H. Krawczyk, B. Wiszniewski
October 2004 **DocEng '04:** Proceedings of the 2004 ACM symposium on Document engineering
Publisher: ACM

Full text available: 🗎 pdf(313.24 KB)

Additional Information: full citation, abstract, references, cited by, index terms, review

This paper describes the lifecycle of a digital historical document, from template-based structure definition through to content extraction from the scanned pages and its final reconstitution as an electronic document (combining content and semantic ...

Keywords: digital libraries, document analysis, document architecture, document engineering, historical documents, text enhancement

2   Digital document life cycle development
Henryk Krawczyk, Bogdan Wiszniewski
September 2003 **ISICT '03:** Proceedings of the 1st international symposium on Information and communication technologies
Publisher: Trinity College Dublin

Full text available: 🗎 pdf(198.83 KB)

Additional Information: full citation, abstract, references

The paper reports on the ongoing project IST-2001-33441-MEMORIAL funded by EU under the Framework 5 Programme, aimed at developing an interactive electronic document model suitable for creating a Virtual Memorial Web portal. Documents are based on personal ...

3   Document digitization lifecycle for complex magazine collection
Sherif Yacoub, John Burns, Paolo Faraboschi, Daniel Ortega, Jose Abad Peiro, Vinay Saxena
November 2005 **DocEng '05:** Proceedings of the 2005 ACM symposium on Document engineering
Publisher: ACM

Full text available: 🗎 pdf(540.79 KB)

Additional Information: full citation, abstract, references, cited by, index terms

The conversion of large collections of documents from paper to digital formats that are suitable for electronic archival is a complex multi-phase process. The creation of good quality images from paper documents is just one phase. To extract relevant ...

Keywords: document analysis and understanding, document digitization, document engineering, preservation of historical content

4 Content publishing framework for interactive paper documents

Moira C. Norrie, Alexios Palinginis, Beat Signer
November 2005 **DocEng '05:** Proceedings of the 2005 ACM symposium on Document engineering
Publisher: ACM

Full text available: .pdf(721.89 KB)   Additional Information: full citation, abstract, references, index terms

Paper persists as an important medium for documents and this has motivated the development of new technologies for interactive paper that enable actions on paper to be linked to digital actions. A major issue that remains is how to integrate these technologies ...

Keywords: interactive paper, publishing framework

5 Extracting mathematical expressions from postscript documents

Michael Yang, Richard Fateman
July 2004 **ISSAC '04:** Proceedings of the 2004 international symposium on Symbolic and algebraic computation
Publisher: ACM

Full text available: .pdf(137.43 KB)   Additional Information: full citation, abstract, references, index terms

Full-text indexing of documents containing mathematics cannot be considered a complete success unless the mathematics symbolism is extracted and represented in a standardized form permitting both searching for formulas, and re-use of this information ...

Keywords: digital library, document image analysis, mathematics, optical character recognition, postscript documents

6 Visual signature based identification of Low-resolution document images

Ardhendu Behera, Denis Lalanne, Rolf Ingold
October 2004 **DocEng '04:** Proceedings of the 2004 ACM symposium on Document engineering
Publisher: ACM

Full text available: .pdf(2.00 MB)   Additional Information: full citation, abstract, references, cited by, index terms

In this paper, we present (a) a method for identifying documents captured from low-resolution devices such as web-cams, digital cameras or mobile phones and (b) a technique for extracting their textual content without performing OCR. The first method ...

Keywords: document visual signature, document-based meeting retrieval, documents' content extraction, low-resolution document image identification

7 Building a test collection for complex document information processing
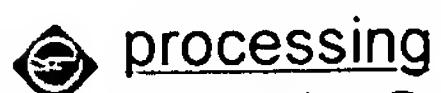
D. Lewis, G. Agam, S. Argamon, O. Frieder, D. Grossman, J. Heard

August 2006 **SIGIR '06:** Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval

Publisher: ACM

Additional Information: full citation, abstract, references, cited by, index terms

Full text available: 📄 pdf(294.83 KB)

Research and development of information access technology for scanned paper documents has been hampered by the lack of public test collections of realistic scope and complexity. As part of a project to create a prototype system for search and mining ...

Keywords: TREC, corpora, metadata, queries, relevance judgments

8 Enhancing composite digital documents using XML-based standoff markup

Peter L. Thomas, David F. Brailsford

November 2005 **DocEng '05:** Proceedings of the 2005 ACM symposium on Document engineering

Publisher: ACM

Full text available: 📄 pdf(695.86 KB)   Additional Information: full citation, abstract, references, index terms

Document representations can rapidly become unwieldy if they try to encapsulate all possible document properties, ranging from abstract structure to detailed rendering and layout.We present a composite document approach wherein an XML-based document ...

Keywords: MathML, MusicXML, PDF, XBL, XML, composite documents, standoff markup

9 Towards smarter documents

Vikas Krishna, Prasad M. Deshpande, Savitha Srinivasan

November 2004 **CIKM '04:** Proceedings of the thirteenth ACM international conference on Information and knowledge management

Publisher: ACM

Additional Information: full citation, abstract, references, cited by, index terms

Full text available: 📄 pdf(224.70 KB)

Document analysis research typically focuses on document image understanding or classic problems in text classification, clustering, summarization and discovery. While that is an important aspect of document management, in practice, documents lifecycles ...

Keywords: classification, content, processes, workflow

10 Structuring documents according to their table of contents

Hervé Déjean, Jean-Luc Meunier

November 2005 **DocEng '05:** Proceedings of the 2005 ACM symposium on Document engineering

Publisher: ACM

Full text available: 📄 pdf(544.22 KB)   Additional Information: full citation, abstract,

In this paper, we present a method for structuring a document according to the information present in its Table of Contents. The detection of the ToC as well as the determination of the parts it refers to in the document body rely on a series of generic ...

Keywords: document structuring, table of contents recognition

11 Logical document conversion: combining functional and formal knowledge
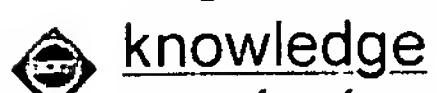
Hervé Déjean, Jean-Luc Meunier
August 2007 **DocEng '07:** Proceedings of the 2007 ACM symposium on Document engineering
Publisher: ACM

Full text available: pdf(297.96 KB)   Additional Information:

We present in this paper a method for document layout analysis based on identifying the function of document elements (what they do). This approach is orthogonal and complementary to the traditional view based on the form of document elements (how they ...

Keywords: combination of knowledge, feedback, functional analysis, logical document analysis, methodology

12 Robust document image understanding technologies

Henry S. Baird, Daniel Lopresti, Brian D. Davison, William M. Pottenger
November 2004 **HDP '04:** Proceedings of the 1st ACM workshop on Hardcopy document processing
Publisher: ACM

Full text available: pdf(92.22 KB)   Additional Information:

No existing document image understanding technology, whether experimental or commercially available, can guarantee high accuracy across the full range of documents of interest to industrial and government agency users. Ideally, users should be able to ...

Keywords: OCR error management, document analysis, information retrieval

13 Supervised learning for the legacy document conversion

Boris Chidlovskii, Jérôme Fuselier
October 2004 **DocEng '04:** Proceedings of the 2004 ACM symposium on Document engineering
Publisher: ACM

Full text available: pdf(180.71 KB)   Additional Information:

We consider the problem of document conversion from the rendering-oriented HTML markup into a semantic-oriented XML annotation defined by user-specific DTDs or XML Schema descriptions. We represent both source and target documents as rooted ordered trees ...

Keywords: XML markup, legacy document conversion, machine learning

14 Extracting relevant named entities for automated expense reimbursement

Guangyu Zhu, Timothy J. Bethea, Vikas Krishna

August 2007 **KDD '07:** Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining

Publisher: ACM

Full text available: pdf(1.16 MB)   Additional Information: full citation, abstract, references, index terms

Expense reimbursement is a time-consuming and labor-intensive process across organizations. In this paper, we present a prototype expense reimbursement system that dramatically reduces the elapsed time and costs involved, by eliminating paper from the ...

Keywords: conditional random fields, document layout analysis, learning, named entity extraction

15 HeinOnline

Richard J. Marisa

January 2001 **JCDL '01:** Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries

Publisher: ACM

Full text available: pdf(217.08 KB)   Additional Information: full citation, abstract, references, index terms

HeinOnline is a new online archive of law journals. Development of He inOnline began in late 1997 through the cooperation of Cornell Information Technologies, William S. Hein & Co., Inc. of Buffalo, NY, and the Cornell Law Library.Built upon the familar ...

Keywords: dienst, digital library, document structure, law journals, metadata, system design

16 UpLib: a universal personal digital library system

William C. Janssen, Kris Popat

November 2003 **DocEng '03:** Proceedings of the 2003 ACM symposium on Document engineering

Publisher: ACM

Additional Information: full citation, abstract, references, cited by, index terms

Full text available: pdf(261.28 KB)

We describe the design and use of a personal digital library system, UpLib. The system consists of a full-text indexed repository accessed through an active agent via a Web interface. It is suitable for personal collections comprising tens of thousands ...

Keywords: document management, document repository, page image, personal digital library, thumbnail interfaces, web interfaces

17 INFTY: an integrated OCR system for mathematical documents

Masakazu Suzuki, Fumikazu Tamari, Ryoji Fukuda, Seiichi Uchida, Toshihiro Kanahori

November 2003 **DocEng '03:** Proceedings of the 2003 ACM symposium on Document engineering

Publisher: ACM

Full text available: 📄 pdf(322.41 KB)

Additional Information: full citation, abstract, references, cited by, index terms

An integrated OCR system for mathematical documents, called INFTY, is presented. INFTY consists of four procedures, i.e., layout analysis, character recognition, structure analysis of mathematical expressions, and manual error correction. In those procedures, ...

Keywords: character and symbol recognition, mathematical OCR, structure analysis of mathematical expressions

18 The processing of digitized works

José Borbinha, João Gil, Gilberto Pedrosa, João Penas
June 2006 **JCDL '06:** Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries
Publisher: ACM

Full text available: 📄 pdf(2.74 MB)

Additional Information: full citation, abstract, references, index terms

This paper describes the processing of digitised works at the National Library of Portugal, as done in the scope of the National Digital Library initiate (BND). This comprises the normalization of the names of the images, the creation of technical metadata, ...

Keywords: METS, OCR, best practices, digital publishing, digitization, image processing, structural metadata

19 A system for understanding imaged infographics and its applications

Weihua Huang, Chew Lim Tan
August 2007 **DocEng '07:** Proceedings of the 2007 ACM symposium on Document engineering
Publisher: ACM

Full text available: 📄 pdf(1.13 MB)

Additional Information: full citation, abstract, references, index terms

Information graphics, or infographics, are visual representations of information, data or knowledge. Understanding of infographics in documents is a relatively new research problem, which becomes more challenging when infographics appear as raster images. ...

Keywords: applications, association of text and graphics, document image understanding, infographics

20 Injecting information into atomic units of text

Yannis Haralambous, Gábor Bella
November 2005 **DocEng '05:** Proceedings of the 2005 ACM symposium on Document engineering
Publisher: ACM

Additional Information: full citation, abstract, references, cited by, index terms

Full text available: 📄 pdf(244.01 KB)

This paper presents a new approach to text processing, based on textemes. These are atomic text units generalising the concepts of character and glyph by merging them in a common data structure, together with an arbitrary number of user-defined properties. ...

Keywords: OpenType, PDF, SVG, Unicode, character, glyph, multilingual typesetting, omega, texteme

Results 1 - 20 of 38   Result page: **1**   2   next   >>

Useful downloads: Adobe Acrobat   QuickTime   Windows Media Player   Real Player

Google™
Patent Search ◐ BETA

| ocr and search documents | Search Patents | Advanced Patent Sea
Google Patent Search |

The "AND" operator is unnecessary -- we include all search terms by default. [details]

Patents

Patents **1 - 10** on **ocr and search documents**. (0.02 seconds)

Sort by relevance | Sort by date (new first) | Sort by date (old first)

Information processing method and apparatus, and storage medium storing ...
US Pat. 6310971 - Filed Jun 28, 1996 - Canon Kabushiki Kaisha
The characters likely to be improperly recognized by the **OCR** are stored in the
... This processing can eliminate **documents** including **search** keys but not ...

System and method for portable document indexing using n-gram word ...
US Pat. 5706365 - Filed Apr 10, 1995 - Rebus Technology, Inc.
However, a detailed index may provide the benefit of reduced **search** 5 times. ...
OCR systems portability of indexed **documents**, are generally very sensitive ...

System and method for searching electronic **documents** created with optical ...
US Pat. 6480838 - Filed Jun 9, 2000
Search engines sometimes fail to locate **documents** that have been created using
scanners and **OCR** software. This is due to the existence of numerous errors in ...

Non-literal textual **search** using fuzzy finite non-deterministic automata
US Pat. 5606690 - Filed Oct 28, 1994 - Canon Inc.
Scanning systems often utilize optical character recognition (**OCR**) that converts
text portions of scanned images into electronic data. Stored **documents** thus ...

Word grouping accuracy value generation
US Pat. 6269188 - Filed Mar 12, 1998 - Canon Kabushiki Kaisha
For example, **OCR** 1 may recognize handwriting particularly accurately. ... to index
the **documents** or provide **search** terms for searching within the document. ...

Non-literal textual **search** using fuzzy finite-state linear non-deterministic ...
US Pat. 6018735 - Filed Aug 22, 1997 - Canon Kabushiki Kaisha
Scanning systems often utilize optical character recognition (**OCR**) that converts
text portions of scanned images into electronic data. Stored **documents** thus ...

Mass document storage and retrieval system
US Pat. 5109439 - Filed Jun 12, 1990
Search in many **OCR** conversion systems. Furthermore, such systems have no provision
... In addition, rather rapid processing of **documents** and storage of the ...

Image matching and retrieval by multi-access redundant hashing
US Pat. 5465353 - Filed Apr 1, 1994 - Ricoh Company, Ltd.
One query that is of interest in the above example of a document database is a
search for **documents** with a given passage of text. ...

System for searching a corpus of document images by user specified document ...
US Pat. 5999664 - Filed Nov 14, 1997 - Xerox Corporation
Many document **search** and retrieval systems rely entirely on the results of applying
OCR (Optical Character Recognition) to every scanned document image. ...

Method and apparatus for providing automated searching and linking of ...

US Pat. 6138129 - Filed Dec 16, 1997 - World One Telecom, Ltd.
using optical character recognition (**OCR**) to create an electronic file(s) of the
one or more **documents**; (2) editing/formatting the **OCR** 30 **documents**; ...

Goooooooooogle ▶

Result Page: **1** 2 3 4 5 6 7 8 9 10   **Next**

| ocr and search documents | | Search Patents |

Google Patent Search Help | Advanced Patent Search

Google Home - About Google - About Google Patent Search

©2007 Google

Google Patent Search ○ SEA

| ocr and search documents and translate keyw | Search Patents | Advanced Patent Sear
Google Patent Search |

The "AND" operator is unnecessary -- we include all search terms by default. [details]

Patents

Patents **1 - 1** on **ocr and search documents and translate keyword**. **(0.02 seconds)**

Sort by relevance | Sort by date (new first) | Sort by date (old first)

Imaged document optical correlation and conversion system

US Pat. 6741743 - Filed Jul 31, 1998 - PRC. Inc.
To **search** for specific ordinary skill will be able to affect various changes,
... of the required to successfully implement the OC and **OCR** aspects „ ™th the ...

| ocr and search documents and translate keyv | Search Patents |

Google Patent Search Help | Advanced Patent Search

Google Home - About Google - About Google Patent Search

©2007 Google

Google™
Patent Search ◯ BETA

| ocr and search documents and translate | | Search Patents |

Advanced Patent Sear
Google Patent Search

The "AND" operator is unnecessary -- we include all search terms by default. [details]

Patents

Patents **1 - 7** on **ocr and search documents and translate**. (**0.02** seconds)

Sort by relevance | Sort by date (new first) | Sort by date (old first)

Imaged document optical correlation and conversion system
US Pat. 6741743 - Filed Jul 31, 1998 - PRC. Inc.
To **search** for specific ordinary skill will be able to affect various changes,
... of the required to successfully implement the OC and **OCR** aspects „ ™th the ...

Document-based query data for information retrieval
US Pat. 6396951 - Filed Dec 23, 1998 - Xerox Corporation
The significant image units are then decoded by optical character recognition (**OCR**)
techniques, and the 45 decoded words can then be used to access ...

Automatic language identification using both N-gram and word information
US Pat. 6167369 - Filed Dec 23, 1998 - Xerox Company
The **search** engine could, for 60 example, automatically identify the language of
a query, **translate** the query into one or more other languages, ...

Apparatus for translation of character codes for application to a data ...
US Pat. 4425626 - Filed Nov 29, 1979 - Honeywell Information Systems Inc.
The character codes read from the **documents** include a number of special characters
... read heads may read **OCR** (Optical Character Recognition) characters. ...

Indexing with translation model for feature regularization
US Pat. 6925436 - Filed Jan 28, 2000 - International Business Machines Corporation
There, the statistical translation model was trained to **translate** from speech
... handwriting indexing, **OCR** indexing) broadens this domain further. ...

System for converting medical information into representative abbreviated ...
US Pat. 5809476 - Filed Mar 23, 1995
... **documents** in which case Optical Character Recognition (**OCR**) could be used.
Alternatively a voice recognition system may be used to directly **translate** ...

Systems and methods for adaptive handwriting recognition
US Pat. 7184591 - Filed May 21, 2003 - Microsoft Corporation
Thus, optical character recognition (**OCR**) technology was developed to utilize.
... the **OCR** software to **translate** a scanned image 65 into editable text. ...

Google Patent Search Help | Advanced Patent Search

Google Home - About Google - About Google Patent Search

©2007 Google

Google™
Patent Search ⊘ BETA

ocr and search documents and output xml	Search Patents	Advanced Patent Sea
		Google Patent Search

The "AND" operator is unnecessary -- we include all search terms by default. [details]

Patents

Patents **1 - 2** on **ocr and search documents and output xml**. (**0.02** seconds)

Sort by relevance | Sort by date (new first) | Sort by date (old first)

User-defined **search** template for extracting information from **documents**
US Pat. 6353840 - Filed Aug 13, 1998 - Ricoh Company, Ltd.
search template. The **documents** are either stored in a storage device such as the
... An optical character recognition unit (**OCR**) 207 converts alphanumeric ...

System and method for automatic preparation of data repositories from ...
US Pat. 6810136 - Filed Oct 18, 2002 - Olive Software Inc.
These time-based definitions typically increase the speed and efficiency of
operation of a **search** engine through the **output** format data, for example. ...

ocr and search documents and output xml	Search Patents

Google Patent Search Help | Advanced Patent Search

Google Home - About Google - About Google Patent Search

©2007 Google